Spatiotemporal Differences of COVID-19 Infection Among Healthcare Workers and Patients in China From January to March 2020

PEIXIAO WANG[1], **XINYAN ZHU**[1,3,4], **WEI GUO**[1], **HUI REN**[1], **AND TAO HU**[2]

[1]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[2]Center for Geographic Analysis, Harvard University, Cambridge, MA 02138, USA
[3]Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China
[4]Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, Wuhan University, Wuhan 430079, China

Corresponding authors: Xinyan Zhu (xinyanzhu@whu.edu.cn) and Tao Hu (taohu@g.harvard.edu)

ABSTRACT Studying the spatiotemporal differences in coronavirus disease (COVID-19) between social groups such as healthcare workers (HCWs) and patients can aid in formulating epidemic containment policies. Most previous studies of the spatiotemporal characteristics of COVID-19 were conducted in a single group and did not explore the differences between groups. To fill this research gap, this study assessed the spatiotemporal characteristics and differences among patients and HCWs infection in Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei). The temporal difference was greater in Wuhan than in the rest of Hubei, and was greater in Hubei (excluding Wuhan) than in the rest of China. The incidence was high in healthcare workers in the early stages of the epidemic. Therefore, it is important to strengthen the protective measures for healthcare workers in the early stage of the epidemic. The spatial difference was less in Wuhan than in the rest of Hubei, and less in Hubei (excluding Wuhan) than in the rest of China. The spatial distribution of healthcare worker infections can be used to infer the spatial distribution of the epidemic in the early stage and to formulate control measures accordingly.

INDEX TERMS Healthcare worker infection, patient infection, spatiotemporal distribution, spatiotemporal differences, COVID-19.

I. INTRODUCTION

The COVID-19 virus was first discovered at the end of 2019, and subsequently spread rapidly worldwide. At present, the world has been heavily hit by the COVID-19 pandemic [1], [2]. Since the outbreak of COVID-19, many researchers have performed classical, epidemiological, mathematical, and statistical analyses to carry out emergency research from the perspectives of pathology [3], [4], epidemiology [5], genomics [6], clinical medicine [7]–[9], and molecular biology [10]. However, when a novel infectious disease is prevalent, relevant preventative measures such as self-isolation, restriction of crowd movement and gathering, and wearing of masks became important emergency containment measures [11]–[13].

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

To achieve targeted and accurate containment, it is necessary to understand the spatiotemporal spread of COVID-19 as time progresses and determine its development trend early-on [14]. In the early stages of the epidemic, patient infection data were a very important data source. Patient infection data contains a wealth of information that can be explored via spatiotemporal analysis [15], [16]. Therefore, many researchers have used this data to explore the spatiotemporal characteristics of COVID-19. However, studies using other data sources to assess the characteristics of the epidemic in social groups (e.g., healthcare workers) are scarce [17]. Studying the spatiotemporal distributions and differences in healthcare worker infections can aid in better formulating epidemic containment policies; however, the spatiotemporal characteristics of COVID-19 between healthcare workers are rarely studied. To fill this research gap, the spatiotemporal characteristics and differences between

healthcare worker infections and patient infections were analyzed from the perspectives of Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei) by combining the data of healthcare workers infection and the data of patients infection. The main contributions of this article are as follows:

(1) The spatiotemporal distributions of healthcare worker infections were analyzed from the perspectives of Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei).

(2) The spatiotemporal differences among healthcare worker infections and patient infections were also analyzed. In addition, the study can provide a scientific reference for relevant departments to formulate epidemic containment policies from a new perspective.

II. RELATED WORKS

Many scientists have researched COVID-19 from different perspectives. In this section, we primarily present a review of relevant studies that explored the spatiotemporal distributions of COVID-19.

To explore the spatiotemporal distribution of the disease is mainly to summarize the spread law of COVID-19 to provide the basis for prevention and control. In the early stage of the epidemic, many studies were conducted based on patient confirmed data. For example, Zhang *et al.* [18] analyzed the epidemiological characteristics of COVID-19 in February 2020 based on patient confirmed data and reported the spatial distribution of COVID-19 early-on in China. Hu *et al.* [19] collected global patient confirmed data from different scales and established a data repository to provide data support for research related to the spatiotemporal distribution of COVID-19. Wang *et al.* [20] used spatiotemporal scanning statistics to detect the hotspots of new cases every week, based on the confirmed cases of COVID-19 at the county level in the United States, thereby characterizing the incidence trend. Further, Zhang *et al.* [21] compared the spatiotemporal characteristics between COVID-19 and SARS at the provincial level, thereby revealing the spread of COVID-19. Sun and Castro [22] analyzed the regional spatiotemporal changes of the epidemic based on the daily new cases in 250 regions of South Korea, thereby evaluating South Korea's containment strategy.

Researchers have also combined patient confirmed data with other data sources to explore the spatiotemporal spread of COVID-19. For example, Zhang *et al.* [23] examined the factors influencing the number of imported cases from Wuhan and the spread speed and pattern of the pandemic by combining national flight and high-speed rail data. Based on mobile phone and confirmed patient data, Jia *et al.* [24] developed a spatiotemporal "risk source" model to determine the geographic distribution and growth pattern of COVID-19 and quickly as well as accurately assess the related risk. Zhang *et al.* [25] used the GeoDetector and the decision tree model to identify the main factors in low- and high-risk areas by combining multi-source data. Loske [26] explored the relationship between COVID-19 spread and transportation volume in food retail logistics by combining transport volume

data and confirmed patient data. Further, Zhang *et al.* [27] measured the imported case risk of COVID-19 from inbound international flights by combining both daily dynamic international air connectivity data and updated global COVID-19 data.

However, the abovementioned studies are mostly based on patient confirmed data. The spatiotemporal distributions of COVID-19 in other groups are rarely explored. Thus, we analyzed the spatiotemporal characteristics of and differences between healthcare worker infections and patient infections from the perspectives of Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei).

III. STUDY AREA AND DATA SOURCES
A. STUDY AREA

As of July 2020, the total number of locally confirmed COVID-19 cases in China had exceeded 80,000. Of these cases, 68,135 were confirmed in Hubei Province, among which 50,340 were confirmed in Wuhan, accounting for 81.5% and 60.80% of the country's cases, respectively. To study the temporal and spatial spread of COVID-19 in China, the study area was divided into three parts: Wuhan City, Hubei Province (excluding Wuhan), and the rest of China. As shown in Figure 1, Hubei Province is in the central region of China, and Wuhan City is in the central region of Hubei Province.

B. DATA SOURCES AND DATA PREPROCESSING
1) DATA SOURCES

The data used in this study were divided into two categories: *Patient Infection Table* and healthcare worker's infection information. The *Patient Infection Table* was mainly obtained from the National Health Commission of the People's Republic of China (http://www.nhc.gov.cn/xcs/yqtb/list_gzbd.shtml) and spanned from January 15, 2020, to June 28, 2020. The confirmed information on healthcare workers was obtained mainly from the Chinese Red Cross Foundation (https://www.crcf.org.cn/article/20672). The Chinese Red Cross Foundation distributes relief funds to every healthcare worker with a confirmed case. As of July 31, 2020, 81 batches of confirmed healthcare workers had received foundation assistance. We first used Python language and web crawler technology to obtain 3,741 reports on confirmed healthcare workers from the Chinese Red Cross Foundation and later used the Baidu API for address matching. After matching their address, details of the city, province, and county of each healthcare worker with a confirmed case were obtained. The data format is shown in Table 1. Each record shows the confirmed time, province, city, and county of each confirmed healthcare worker.

2) DATA PREPROCESSING

The *Patient Infection Table* is collected and processed by a team at the China Data Center and is shared on the dataverse
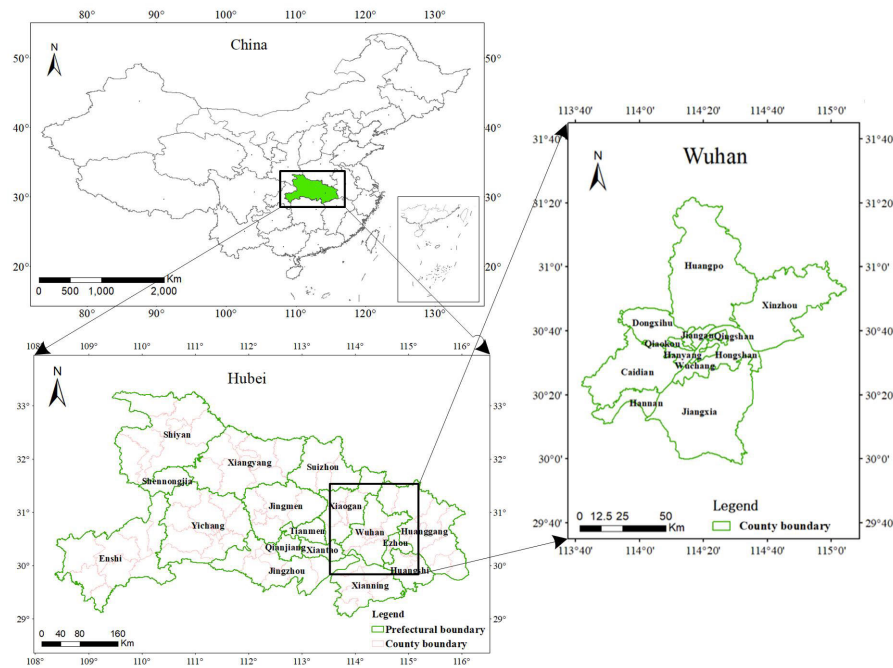
FIGURE 1. Sketch map of the study area.

TABLE 1. Sample of healthcare worker infection data.

User ID	Date	Province	City	County
1	2020-01-15	Hubei	Wuhan	Huangpo
2	2020-01-15	Hubei	Jingmen	Zhongxiang
3	2020-02-04	Shandong	Qingdao	Shinan
...	...	...	...	...
3741	2020-02-01	Beijing	Beijing	Xicheng

platform of Harvard University [19], [28]. Therefore, the *Patient Infection Table* does not require additional data pre-processing. In this study, we mainly conducted data pre-processing for the confirmed case information of healthcare workers.

Infection information on healthcare workers is mainly reported to the China Red Cross Foundation in two ways [29]: (1) direct reporting by individuals to the Red Cross Foundation of China, and (2) collection by the person's respective hospital which then reports the information to the Red Cross Foundation of China. After examination and approval by the Red Cross Foundation of China, the rescue information is published on the website, after which the hospital may review it. Those who fail to pass the review do not qualify to be rescued. Therefore, data records of unqualified persons should be deleted from the original data. Second, the Chinese Red Cross Foundation not only rescued the infected healthcare worker, but also the infected or diseased staff during the epidemic; such data records need to be deleted. As shown in Figure 2, after data preprocessing, the data of a total of 3,701 healthcare workers infected with COVID-19 remained, including 3,667 from Hubei Province and 3,060 from Wuhan City.
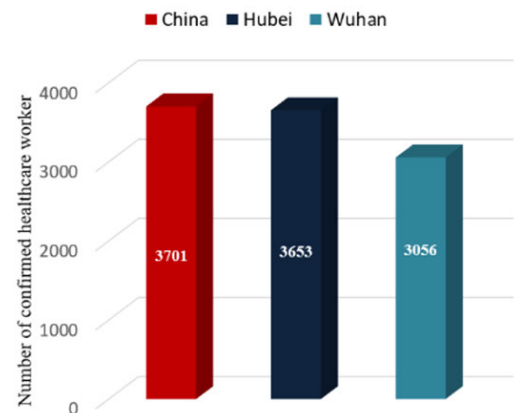

FIGURE 2. Characteristics of healthcare worker infection data.

IV. METHODS

The overall framework of the study is shown in Figure 3. First, the *Healthcare Worker Infection Tables* in China, Hubei, and Wuhan, which recorded a high number of healthcare workers infection each day, were constructed. Second, based on the *Infection Tables*, relevant statistical indicators such as the mean, variance, change speed (rise and decline periods), and the inflection point were used to analyze the time spread characteristics of healthcare worker and patient infections; we used Pearson's correlation coefficient to calculate the time correlation between healthcare worker infection and patient infection. Finally, the thematic map method was used to analyze the spatial spread characteristics of infections among healthcare workers and patients. We also calculated the spatial correlation between healthcare worker and patient infections.
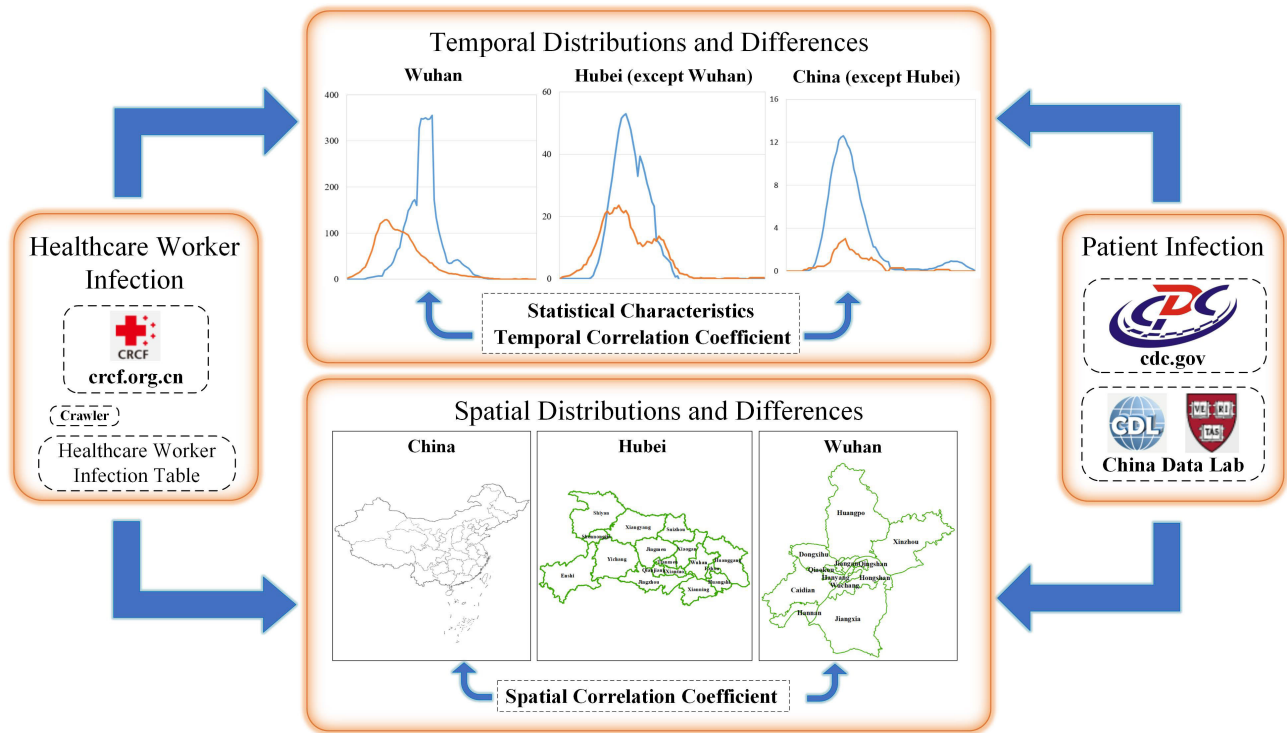
FIGURE 3. Technical framework of spatiotemporal distribution and differences in healthcare workers and patient infection.

A. CONSTRUCTION OF THE HEALTHCARE WORKER INFECTION TABLE

The healthcare worker infection data which were collected using the crawler record the basic information of confirmed healthcare workers, however, the data cannot visually display the temporal and spatial distributions and spread of the confirmed healthcare worker's infections. Therefore, based on the modeling idea of "province - city - county," this study constructed *Healthcare Worker Infection Tables* based on the Partition Statistics from three perspectives: China, Hubei, and Wuhan.

The data structure for the *Healthcare Worker Infection Tables* are shown in Figure 4. The *Table(S, T)* represents a spatiotemporal dataset, where S and T are the spatial and temporal dimensions, respectively; $S = \{s_1, s_2, \ldots, s_m\}$ represents the row index of the table, and $T = \{t_1, t_2, \ldots, t_n\}$ represents the column index of the table.

Taking the *Healthcare Worker Infection Table* of Wuhan as an example, the construction process is shown in Algorithm 1. $\{cms_i\}$ represents healthcare worker infection data, and $Table^{Wuhan}$ represents the *Healthcare Worker Infection Table* from the perspective of Wuhan. First, $Table^{Wuhan}$ is initialized to 0 (line 1). Then, the values in the table are assigned iteratively (lines 2-4), and $Table^{Wuhan}$ describes in detail the changes in healthcare worker infections in every county of Wuhan over time. Final, $Table^{Wuhan}$ is returned (line 5). The corresponding *Healthcare Worker Infection Tables* in China and Hubei are expressed as $Table^{China}$ and $Table^{Hubei}$, respectively. The Healthcare Worker Infection Table of China

records the number of healthcare worker infections every day in each province of China, and the *Healthcare Worker Infection Table* of Hubei records the number of confirmed healthcare worker infections every day in each city of Hubei Province.

Algorithm 1 Construction of Infection Table

Require: *Healthcare Infection Data* $= \{cms_i\}$
Ensure: $Table^{Wuhan}(S, T)$
1: $Table^{Wuhan}(S, T) = 0$
2: **for** each $cms \in \{cms_i\}$ **do**
3:     $Table^{Wuhan}(cms.County, cms.Date) += 1$
4: **end for**
5: **return** $Table^{Wuhan}$

B. TEMPORAL AND SPATIAL CHARACTERISTICS INDEX

The *Healthcare Worker Infection Table* and the *Patient Infection Table* are essentially a collection of time series that contain spatial location information. Therefore, we further analyzed the temporal and spatial distribution characteristics of healthcare worker infections and patient infections.

In terms of temporal characteristics, we used eight indicators, namely the mean, standard variance, peak, rising period, rising rate, falling period, falling rate, and inflection point, to study the temporal characteristics of healthcare worker and patient infections in China (excluding Hubei), Hubei (excluding Wuhan), and Wuhan. In the Chinese scale, the main reasons for excluding Hubei Province are as follows:
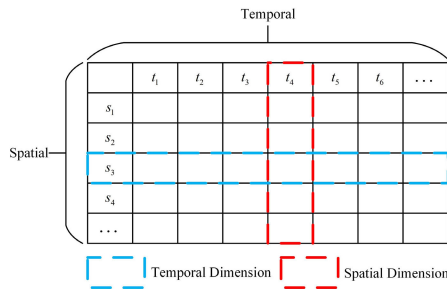
compared with other provinces, the number of patients and healthcare worker infections in Hubei Province is extremely high; if Hubei Province is not excluded, the temporal characteristics of China are only determined by the data of Hubei. Similarly, in the Chinese scale, Wuhan city is also excluded.

In terms of spatial characteristics, we divided the *Healthcare Worker Infection Table* and the *Patient Infection Table* based on time and used thematic maps to show the spatial distribution characteristics of healthcare worker and patient infections at each time node from China, Hubei, and Wuhan.

We determined the spatiotemporal differences between infections in healthcare workers and patients by assessing the differences in indices. A relatively small difference in indices is indicative of a relatively small spatiotemporal difference between infections in healthcare workers and patients, and vice versa.

C. PEARSON CORRELATION COEFFICIENT

The Pearson correlation coefficient [30] was calculated to quantitatively measure the spatiotemporal correlation between healthcare worker infections and patient infections, i.e., the difference in space and time between healthcare worker infections and patient infections. As shown in Figure 5, the spatial correlation was obtained by t_3^h and t_3^p, and the temporal correlation was obtained by s_3^h and s_3^p [31]. The temporal and spatial correlation coefficients of the two series were calculated using Equations (1) and (2), respectively:

$$r_{temporal} = \frac{Cov(s_i^p, s_i^h)}{\sqrt{D(s_i^p)}\sqrt{D(s_i^h)}} \quad (1)$$

$$r_{spatial} = \frac{Cov(t_i^p, t_i^h)}{\sqrt{D(t_i^p)}\sqrt{D(t_i^h)}} \quad (2)$$

where $Cov(X, Y)$ is the covariance of random variables X and Y, $D(X)$ is the variance of the random variable X, and $\sqrt{D(X)}$ is the standard deviation of the random variable X. The value range of correlation coefficient r is $[-1, 1]$. When $r > 0$, the random variables X and Y are positive correlations, if $r = 1$, random variables X and Y are completely positively correlated; when $r < 0$, the random variables X and Y are negatively correlated, if $r = -1$, then the random variables X and Y are completely negatively correlated. To measure the time correlation between healthcare worker infections

and patient infections, X and Y, respectively, represent the time series of healthcare worker infections and patient infections in the same study object. Similarly, to measure the spatial correlation between healthcare workers and patients, X and Y, respectively, represent the number of healthcare worker infections and patient infections in multiple regions at a specific time. A strong positive correlation between two random variables is indicative of a relatively small spatiotemporal difference between infections in healthcare workers and patients, and vice versa.

V. EXPERIMENTAL RESULTS AND ANALYSIS
A. TEMPORAL DIFFERENCES BETWEEN HEALTHCARE WORKER AND PATIENT INFECTIONS

In this study, the statistical graph was first used to qualitatively describe the changes in the number of daily healthcare worker infections and patient infections over time. To make the curve smoother, we used the moving average method using a time window of seven days. The results are shown in Figure 6. Overall, the curves of daily healthcare workers infection and daily patients infection in China (excluding Hubei), Hubei Province (excluding Wuhan), and Wuhan City all initially showed a rising trend followed by a falling one. At the same time, the peaks of infections in China (excluding Hubei), Hubei (excluding Wuhan), and Wuhan have successively decreased. However, the curves of confirmed healthcare workers and confirmed patients were significantly different over time. For example, Figure 6(a) shows the onset of infection in healthcare workers, while it was absent in patients during the first week. There are two main reasons for this difference. First, the statistical approaches of the two data sources are different, which leads to the difference between the data. Second, and more importantly, data of healthcare worker infections were retrospective, while patient infections data were reported in real time.

In addition, there were also significant differences in the statistical characteristics of the curves for healthcare worker and patient infections (Table 2). For example, The mean times of onset of infection for healthcare workers in Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei) were January 27, February 1, and February 2, 2020, while those for patients were February 8, February 2, and January 31, 2020, the differences being 12, 1, and 2 days, respectively. The standard deviation values of infection onset time for healthcare workers in Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei) were 9, 10, and 6 days, while those for patients were 7, 7, and 6 days, the differences being 2, 3, and 0 days, respectively. The results showed the mean time of onset for healthcare workers infection in Wuhan and Hubei (except Wuhan) was earlier than those for patients infection, and the mean times of onset for healthcare workers infection in China (except Hubei) was later than those for patients infection. There are two main reasons. First, the epidemics in Wuhan and Hubei (except Wuhan) are more serious. Compared with patient infections, healthcare worker infections are
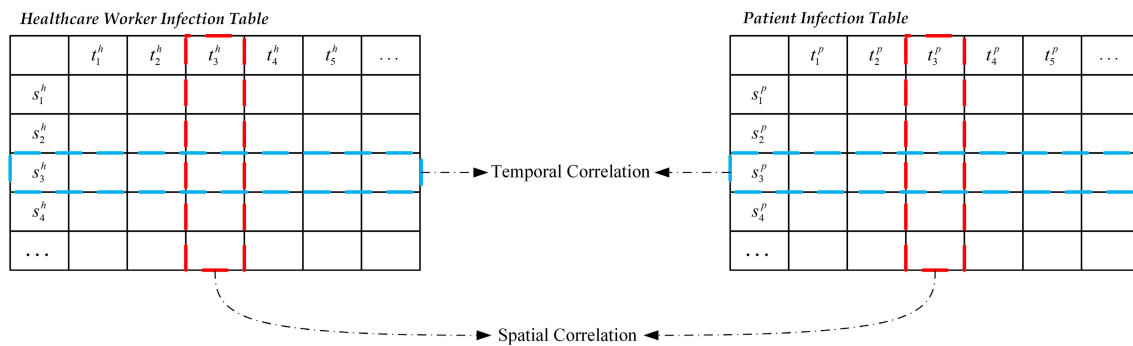
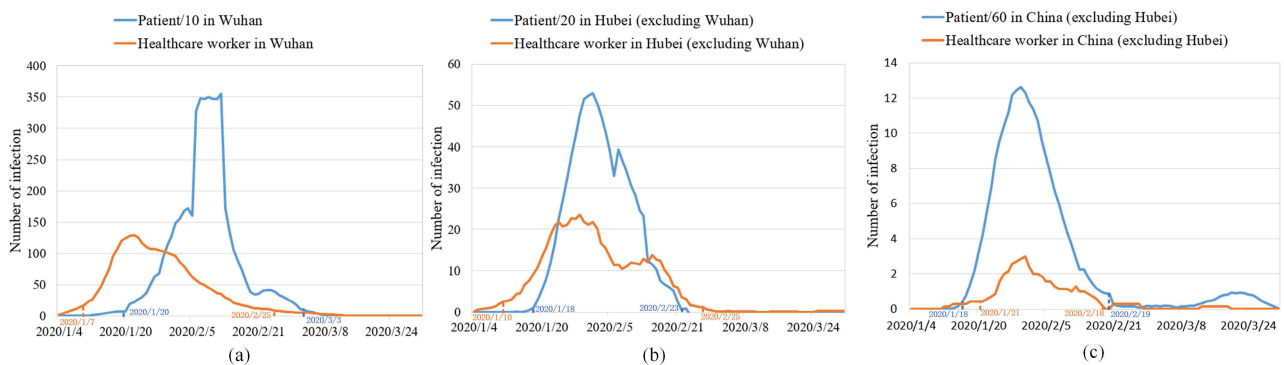**FIGURE 5.** Temporal and spatial correlation coefficients.

FIGURE 6. Comparison between the temporal characteristics of healthcare worker infection data and patient infection data: (a) Wuhan City, (b) Hubei Province excluding Wuhan, and (c) China excluding Hubei.

more likely to be detected. Second, healthcare workers have professional protective measures, so it is easier to control the spread of the epidemic among healthcare workers.

In addition, Pearson's correlation coefficient was also used to measure the temporal correlation between healthcare worker and patient infections (Figure 7). The correlation coefficient r values for Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei) were 0.26, 0.852, and 0.95, respectively. There are two main reasons for the lowest temporal correlation coefficient r in Wuhan. First, in the early stage of the epidemic, the protection of healthcare workers was insufficient, which made it easy to cause large-scale infections in hospitals. Second, compared with patient infections, infections in healthcare workers are more easily detected, leading to them being counted. The correlation coefficient also confirms a temporal difference between healthcare worker and patient infections.

In general, healthcare worker infection is contained earlier and faster. The temporal difference in Wuhan was greater than that in Hubei (excluding Wuhan), and the temporal difference in Hubei (excluding Wuhan) was greater than that in China (excluding Hubei). Combined with the time of epidemic occurrence in different regions, the temporal difference in the early phase was greater than that in the later phase, and the temporal difference in the early-onset region was greater than that in the later-onset region.

TABLE 2. Statistical characteristics.

Index	Study area	Healthcare Worker	Patient
	Wuhan	Jan 27	Feb 8
Mean	Hubei(outside Wuhan)	Feb 1	Feb 2
	China(outside Wuhan)	Feb 2	Jan 31
Standard deviation	Wuhan	9 days	7 days
	Hubei(outside Wuhan)	10 days	7 days
	China(outside Wuhan)	6 days	6 days
Inflection point	Wuhan	Jan 22	Feb 12
	Hubei(outside Wuhan)	Jan 29	Feb 1
	China(outside Wuhan)	Jan 30	Jan 31
Peak	Wuhan	128	3550
	Hubei(outside Wuhan)	23	1058
	China(outside Wuhan)	3	756
Rising period	Wuhan	15 days	23 days
	Hubei(outside Wuhan)	19 days	14 days
	China(outside Wuhan)	12 days	10 days
Rising rate	Wuhan	64	1617
	Hubei(outside Wuhan)	12	550
	China(outside Wuhan)	1	464
Falling rate	Wuhan	34 days	20 days
	Hubei(outside Wuhan)	27 days	22 days
	China(outside Wuhan)	19 days	19 days
Falling rate	Wuhan	55	525
	Hubei(outside Wuhan)	11	450
	China(outside Wuhan)	1	334

B. SPATIAL DIFFERENCES BETWEEN HEALTHCARE WORKER INFECTION AND PATIENT INFECTION

The spread of COVID-19 in space is a dynamic process. To further compare the spatial distribution characteristics and differences between healthcare worker and patient infections,
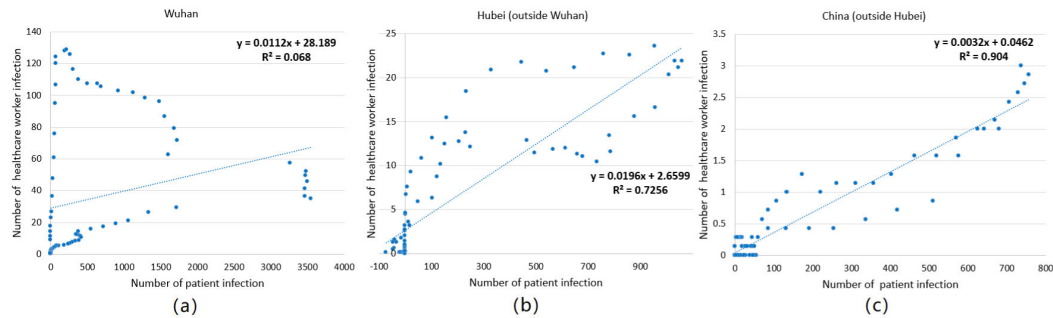
FIGURE 7. Temporal correlations of healthcare worker infection data and patient infection data: (a) Wuhan City, (b) Hubei Province excluding Wuhan, and (c) China excluding Hubei.

we first analyzed the spatial distribution characteristics of healthcare worker and patient infections at each time node with a time interval of 10 days and then quantitatively determined the spatial correlation between healthcare worker and patient infections at each time node. In this section, the period from January 7, 2020, to February 16, 2020 was selected as the research interval.

1) SPATIAL CHARACTERISTICS OF HEALTHCARE WORKER INFECTION

Figure 8 shows the spatial distribution of healthcare worker infections in China, Hubei, and Wuhan. In China, healthcare worker infections first occurred in Hubei, and then the phenomenon of healthcare worker infection gradually spread outwards through Hubei. As of February 6, 2020, the number of healthcare worker infections in different provinces of China significantly differed. Generally, the provinces with many healthcare worker infections in China were mainly located in the southeast coastal areas of China; these provinces are close to Hubei Province and have better economic development. The provinces with relatively low healthcare worker infections were mainly located in northwestern China; these provinces are far from Hubei and economically underdeveloped.

In Hubei, healthcare worker infections first occurred in Wuhan and Huanggang and then spread across, with Wuhan and Huanggang essentially being the center. As of January 27, 2020, most of the cities in Hubei Province had different degrees of healthcare worker infections, among which Wuhan and Huanggang had more more severe infections. Generally, cities with more severe healthcare worker infections are mainly concentrated in the eastern part of Hubei province, while the cities with fewer healthcare worker infections are mainly concentrated in the western part of Hubei Province.

In Wuhan, healthcare worker infections first occurred in Wuhan Jianghan District and Jiangan District and then spread to the two adjacent districts. As of February 06, 2020, all districts in Wuhan had different degrees of healthcare worker infection, among which the districts with more severe healthcare worker infections mainly occurred in the central district of Wuhan, such as Jianghan and Wuchang, where the number

of healthcare worker infections exceeded 500. The districts with mild healthcare worker infections were mainly located in Jiangxia and Huangpo outside Wuhan, but there were still more than 10 cases of healthcare worker infections. In general, healthcare worker infections in Wuhan were severe.

2) SPATIAL CHARACTERISTICS OF PATIENT INFECTION

As a comparison with the spatial distribution characteristics of healthcare infection, we showed the spatial distribution characteristics of patient infection during the same period. As it is difficult to obtain data on patient infections at the county level in Wuhan from mid-January to mid-February, we showed the spatial distribution characteristics of patient infection only at the scale of China and Hubei. The results are shown in Figure 9.

In China, the first case of infection in a patient was reported in Hubei and then spread outward through Hubei. As of February 06, 2020, the numbers of patient infections in different provinces of China were significantly different. Compared with healthcare worker infections, in addition to the difference in the number of infections, there were also certain differences in the spatial distribution of infection. For example, the patient infections in Guangdong and Zhejiang were relatively serious, but the healthcare infection in the two provinces were relatively mild. This may be attributable to the differences in the epidemic prevention strategies of different provinces. For example, Guangdong Province acquired more valuable experience during the SARS outbreak and thus managed to swiftly protect healthcare workers.

In Hubei, the first case of infection in a patient was reported in Wuhan City and then spread outward throughout Wuhan City. As of February 06, 2020, the spatial difference was relatively small, except for the large difference in the number of infections. In other words, cities with relatively more severe patient infections have relatively more severe healthcare infections, and cities with less patient infections have less healthcare infections.

Further, to measure the degree of correlation quantitatively between healthcare worker infection and patient infection, we analyzed the spatial correlation between healthcare worker and patient infections (Table 3). The correlation
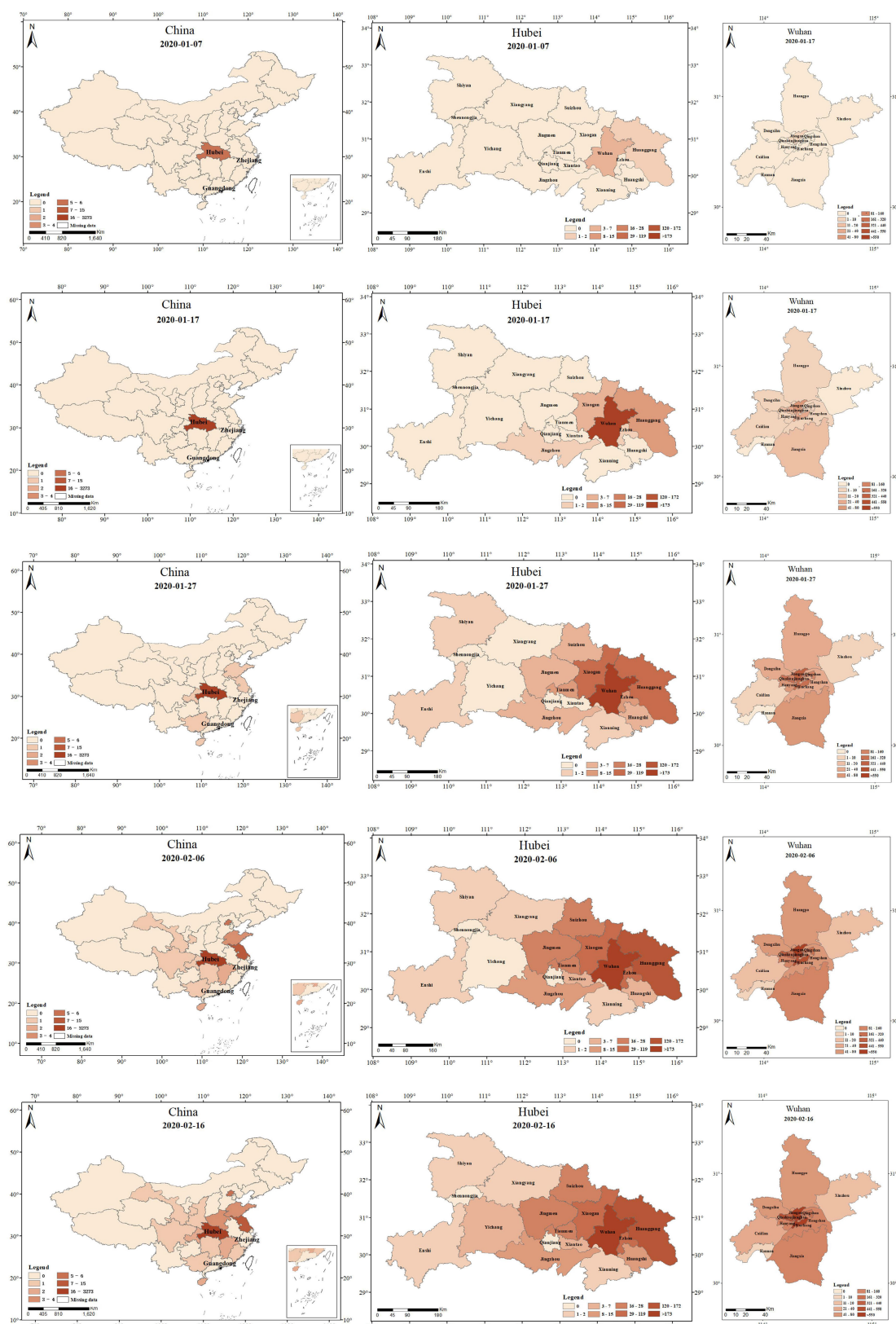
FIGURE 8. Spatial distribution characteristics of healthcare worker infections at different time nodes in China, Hubei, and Wuhan in 2020: (a) on January 7, (b) on January 17, (c) on January 27, (d) on February 6, (e) on February 16.
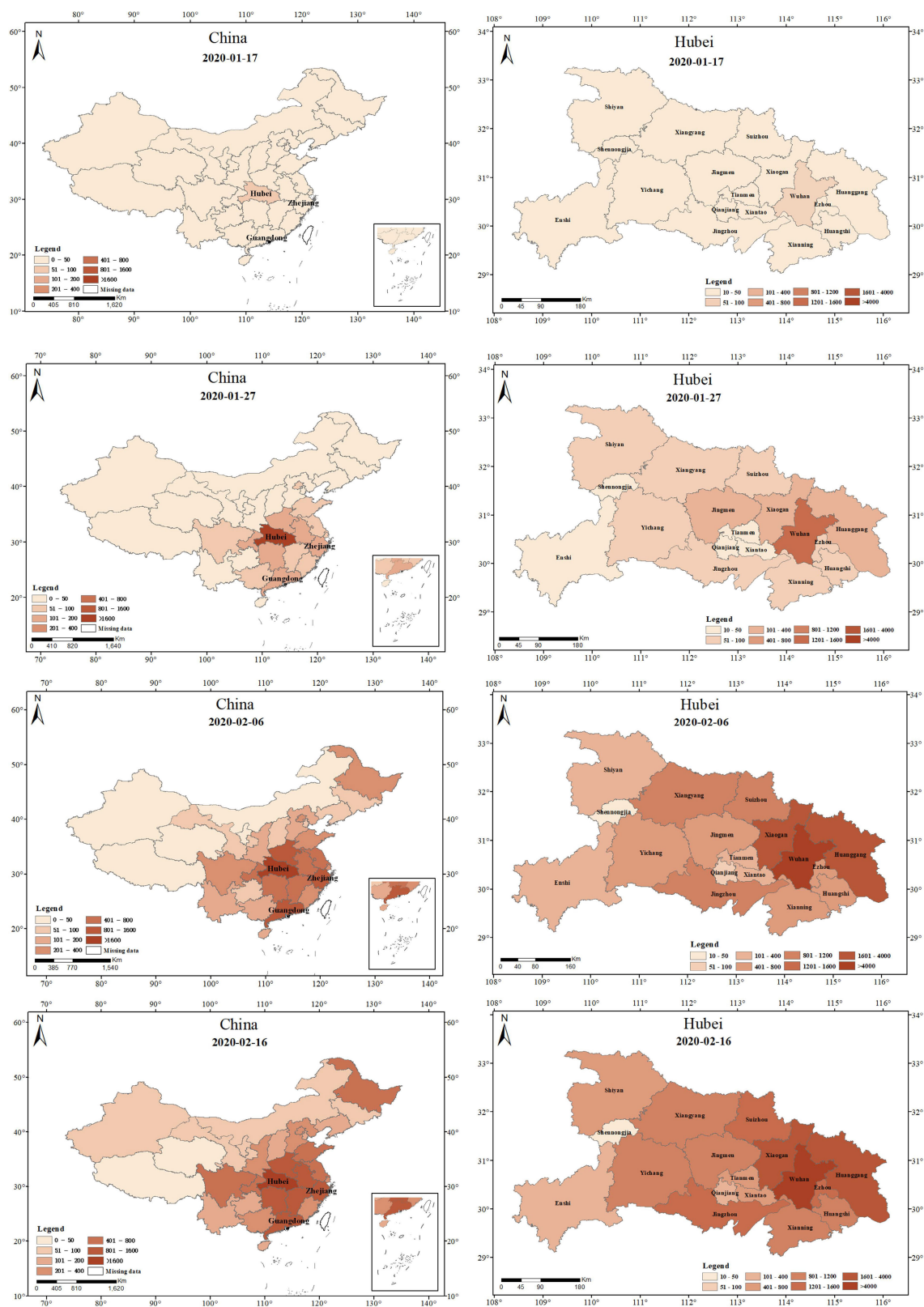
FIGURE 9. Spatial distribution characteristics of patient infections at different time nodes in China and Hubei in 2020: (a) on January 17, (b) on January 27, (c) on February 6, (d) on February 16.

Date	China (outside Hubei)	Hubei (outside Wuhan)
2020-01-27	$r = 0.132$	$r = 0.845$
2020-02-06	$r = 0.159$	$r = 0.816$
2020-02-16	$r = 0.252$	$r = 0.859$

coefficients, r, for China (except Hubei) on January 27, February 06, and February 16 were 0.132, 0.135, and 0.252 respectively, indicating that the spatial distributions of healthcare worker and patient infections in China (except Hubei) are quite different. The correlation coefficients, r, for Hubei (except Wuhan) on January 27, February 06, and February 16 were 0.845, 0.816, and 0.859, respectively, which indicated that the spatial distribution difference between healthcare worker and patient infections in Hubei (except Wuhan) was small.

In general, in addition to the differences in the number of healthcare worker and patient infections, the spatial distribution was different to some extent. The spatial distribution difference in Hubei (except Wuhan) was greater than that in China (except Hubei), that is, the difference in the late stage was greater than that in the early stage; the difference in late-onset regions was greater than that in the early-onset regions.

C. LIMITATIONS OF THE STUDY

The limitations of this study are as follows: (1) the coverage of healthcare worker infection data was narrow, and this study only examined the temporal and spatial distributions and differences between healthcare worker and patient infections in China. Owing to the difficulty in obtaining data on healthcare worker infections in other countries, the temporal and spatial distributions and differences in other countries were not compared and analyzed. (2) The infection data of healthcare workers were not comprehensive. As the information on healthcare worker infections is published on the official website of the Red Cross Foundation in batches, information may continue to be published in the future; thus, this study may be missing information. (3) In terms of the spatial distribution, it is difficult to obtain patient infection data at the county level in Wuhan for the same period. This study only examined the spatial distributions and differences in China (excluding Hubei) and Hubei (excluding Wuhan). In response to the above problems, future research direction will focus on further collecting domestic and foreign healthcare workers and patient infection data to analyze the temporal and spatial distributions and differences more accurately and comprehensively between healthcare worker and patient infections.

VI. DISCUSSIONS AND CONCLUSION

At present, COVID-19 is still a pandemic. As specific vaccines and fundamental treatments take a long time, relevant measures such as self-isolation, restriction of crowd activities and gathering, and wearing of masks have become important emergency containment measures. To accurately formulate

epidemic containment policies, it is important to study the temporal and spatial spread of COVID-19 in social groups.

This study analyzed the temporal and spatial spread characteristics of and differences in COVID-19 among patients and healthcare workers from three perspectives: China (excluding Hubei), Hubei (excluding Wuhan), and Wuhan.

In terms of temporal characteristics, healthcare worker and patient infections in Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei) initially tended to increase and then decrease; the peak value of infection gradually decreased. However, there are great differences between the two curves. For example, the correlation coefficients, r, between healthcare worker and patient infections for Wuhan, Hubei (excluding Wuhan), and China (excluding Hubei) were 0.26, 0.852, and 0.95, respectively. In other words, the temporal differences between healthcare worker and patient infections were relatively large in Wuhan, whereas they decreased successively in Hubei (excluding Wuhan) and China (excluding Hubei). Combined with the time of the occurrence of the epidemic in different regions, the temporal difference in the early stage is greater than that in the later stage between healthcare worker and patient infections, and the temporal difference in the early-onset region is greater than that in the late-onset region. Through temporal difference, different policies can be formulated for different social groups through the following aspects. For example, it is very important to strengthen the protective measures for healthcare workers in the early stage of the epidemic. Because healthcare workers work in a relatively closed space, the battle is on the frontline of the epidemic. Even if the number of patients infected is small, it can still result in many infections among healthcare workers. In addition, we also can use the temporal difference to predict the peak time of infection in patients [32].

In terms of spatial characteristics, healthcare worker and patient infections in China spread mainly through Hubei as the epicenter. Healthcare worker and patient infections in Hubei spread mainly through Wuhan and Huanggang as the epicenter. Healthcare worker infection and patient infections in Wuhan spread mainly through Jianghan and Jiangan as the epicenter. In addition, the spatial distributions of healthcare worker and patient infections also differed from different scales. For example, on January 27, February 06, and February 16, 2020, the correlation coefficients, r, between healthcare worker infection and patient infection for China (excluding Hubei) were 0.132, 0.135, and 0.252, respectively. Further, the correlation coefficients, r, between healthcare worker infection and patient infection in Hubei (excluding Wuhan) were 0.845, 0.816, and 0.859, respectively. Combined with the time of epidemic occurrence in different regions, the spatial difference between healthcare worker infection and patient infections in the early stage was less than that in the later stage, and the spatial difference in the early-onset region was less than that in the late-onset region between healthcare worker and patient infections. The results of this study offer significant guidance for epidemic containment departments to accurately formulate different epidemic

containment policies for different social groups. Through spatial correlation, we can use the spatial distribution of healthcare worker infection to infer the spatial distribution of patient infection in the early stage of the epidemic.

REFERENCES

[1] C. Biscayart, P. Angeleri, S. Lloveras, T. D. S. S. Chaves, P. Schlagenhauf, and A. J. Rodríguez-Morales, "The next big threat to global health? 2019 novel coronavirus (2019-nCoV): What advice can we give to travellers?— Interim recommendations january 2020, from the Latin-American society for travel medicine (SLAMVI)," *Travel Med. Infectious Disease*, vol. 33, Jan. 2020, Art. no. 101567.

[2] Q. Liu, D. Sha, W. Liu, P. Houser, L. Zhang, R. Hou, H. Lan, C. Flynn, M. Lu, T. Hu, and C. Yang, "Spatiotemporal patterns of COVID-19 impact on human activities and environment in mainland China using night-time light and air quality data," *Remote Sens.*, vol. 12, no. 10, p. 1576, May 2020.

[3] M. Ackermann, S. E. Verleden, M. Kuehnel, A. Haverich, T. Welte, F. Laenger, A. Vanstapel, C. Werlein, H. Stark, A. Tzankov, W. W. Li, V. W. Li, S. J. Mentzer, and D. Jonigk, "Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in covid-19," *New England J. Med.*, vol. 383, no. 2, pp. 120–128, Jul. 2020.

[4] P. Xie, W. Ma, H. Tang, and D. Liu, "Severe COVID-19: A review of recent progress with a look toward the future," *Frontiers Public Health*, vol. 8, p. 189, May 2020.

[5] L. Fu, B. Wang, T. Yuan, X. Chen, Y. Ao, T. Fitzpatrick, P. Li, Y. Zhou, Y. F. Lin, Q. Duan, and G. Luo, "Clinical characteristics of coronavirus disease 2019 (COVID-19) in China: A systematic review and meta-analysis," *J. Infection*, vol. 80, no. 6, pp. 656–665, 2020.

[6] P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, and H. D. Chen, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, Mar. 2020.

[7] Z. Fang *et al.*, "Clinical characteristics of coronavirus disease 2019 (COVID-19): An updated systematic review," *medRxiv*, 2020, doi: 10.1101/2020.03.07.20032573.

[8] W. Guan, Z. Y. Ni, Y. Hu, W. H. Liang, C. Q. Ou, J. X. He, L. Liu, H. Shan, C. L. Lei, D. S. Hui, and B. Du, "Clinical characteristics of coronavirus disease 2019 in China," *New England J. Med.*, vol. 382, pp. 1708–1720, Apr. 2020.

[9] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, and Z. Cheng, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, pp. 497–506, May 2020.

[10] A. Pal, A. Pawar, K. Goswami, P. Sharma, and R. Prasad, "Hydroxychloroquine and COVID-19: A cellular and molecular biology based update," *Indian J. Clin. Biochem.*, vol. 35, no. 3, pp. 274–284, 2020.

[11] I. Franch-Pardo, B. M. Napoletano, F. Rosete-Verges, and L. Billa, "Spatial analysis and GIS in the study of COVID-19. A review," *Sci. Total Environ.*, vol. 739, Oct. 2020, Art. no. 140033.

[12] S. Bialek, V. Bowen, N. Chow, A. Curns, R. Gierke, A. Hall, M. Hughes, T. Pilishvili, M. Ritchey, K. Roguski, B. Silk, T. Skoff, P. Sundararaman, E. Ussery, M. Vasser, H. Whitham, J. Wen, and C. COVID-19 Response Team, "Geographic differences in COVID-19 cases, deaths, and incidence—United states, february 12–April 7, 2020," *MMWR. Morbidity Mortality Weekly Rep.*, vol. 69, no. 15, pp. 465–471, Apr. 2020.

[13] B. Nussbaumer-Streit *et al.*, "Quarantine alone or in combination with other public health measures to control COVID-19: A rapid review," *Cochrane Database Syst. Rev.*, no. 9, 2020, doi: 10.1002/14651858.CD013574.pub2.

[14] S. Chen, Q. Li, S. Gao, Y. Kang, and X. Shi, "Mitigating COVID-19 outbreak via high testing capacity and strong transmission-intervention in the United States," *medRxiv*, 2020, doi: 10.1101/2020.04.03.20052720.

[15] X. Zhu, T. Hu, X. Ye, W. Guo, L. Huang, H. Xiong, H. Ai, B. She, Q. Xiong, and L. Duan, "Development and implementation of a dynamic and 4D GIS based on semantic location model," *Geo-Spatial Inf. Sci.*, vol. 22, no. 3, pp. 193–213, Jul. 2019.

[16] B. She, T. Hu, X. Zhu, and S. Bao, "Bridging open source tools and geoportals for interactive spatial data analytics," *Geo-Spatial Inf. Sci.*, vol. 22, no. 3, pp. 185–192, Jul. 2019.

[17] J. Wang, M. Zhou, and F. Liu, "Reasons for healthcare workers becoming infected with novel coronavirus disease 2019 (COVID-19) in China," *J. Hospital Infection*, vol. 105, no. 1, pp. 100–101, May 2020.

[18] C. P. E. R. E. Novel, "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China," *Zhonghua Liu Xing Bing Xue Za Zhi= Zhonghua Liuxingbingxue Zazhi*, vol. 41, no. 2, p. 145, 2020.

[19] T. Hu, W. W. Guan, X. Zhu, Y. Shao, L. Liu, J. Du, H. Liu, H. Zhou, J. Wang, B. She, and L. Zhang, "Building an open resources repository for COVID-19 research," *Data Inf. Manage.*, vol. 4, no. 3, pp. 1–18, 2020.

[20] Y. Wang, Y. Liu, J. Struthers, and M. Lian, "Spatiotemporal characteristics of the COVID-19 epidemic in the United States," *Clin. Infectious Diseases*, 2020, doi: 10.1093/cid/ciaa934.

[21] X. Zhang, H. Rao, Y. Wu, Y. Huang, and H. Dai, "Comparison of spatiotemporal characteristics of the COVID-19 and SARS outbreaks in mainland China," *BMC Infectious Diseases*, vol. 20, no. 1, p. 805, 2020, doi: 10.1186/s12879-020-05537-y.

[22] S. Kim and M. C. Castro, "Spatiotemporal pattern of COVID-19 and government response in South Korea (as of may 31, 2020)," *Int. J. Infectious Diseases*, vol. 98, pp. 328–333, Sep. 2020.

[23] Y. Zhang, A. Zhang, and J. Wang, "Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China," *Transp. Policy*, vol. 94, pp. 34–42, Aug. 2020.

[24] J. S. Jia, X. Lu, Y. Yuan, G. Xu, J. Jia, and N. A. Christakis, "Population flow drives Spatio-temporal distribution of COVID-19 in China," *Nature*, vol. 582, pp. 389–394, Jun. 2020.

[25] Y. Zhang, Y. Li, B. Yang, X. Zheng, and M. Chen, "Risk assessment of COVID-19 based on multisource data from a geographical viewpoint," *IEEE Access*, vol. 8, pp. 125702–125713, 2020.

[26] D. Loske, "The impact of COVID-19 on transport volume and freight capacity dynamics: An empirical analysis in german food retail logistics," *Transp. Res. Interdiscipl. Perspect.*, vol. 6, Jul. 2020, Art. no. 100165.

[27] L. Zhang, H. Yang, K. Wang, Y. Zhan, and L. Bian, "Measuring imported case risk of COVID-19 from inbound international flights—A case study on China," *J. Air Transp. Manage.*, vol. 89, Oct. 2020, Art. no. 101918.

[28] C. Yang, D. Sha, Q. Liu, Y. Li, H. Lan, W. W. Guan, T. Hu, Z. Li, Z. Zhang, J. H. Thompson, and Z. Wang, "Taking the pulse of COVID-19: A spatiotemporal perspective," *Int. J. Digit. Earth*, vol. 13, no. 10, pp. 1186–1211, 2020.

[29] C. R. C. Foundation. (Jul. 2020). *Management Rules of Chinese Red Cross Foundation Bytedance Healthcare Workers Humanitarian Relief Fund (Revised Edition)*. [Online]. Available: https://new.crcf.org.cn/article/19718

[30] I. Cohen, Y. Huang, J. Chen, and J. Benesty, "Pearson correlation coefficient," in *Noise Reduction in Speech Processing*. Berlin, Germany: Springer, 2009, ch. 5, pp. 1–4, doi: 10.1007/978-3-642-00296-0.

[31] S. Cheng and F. Lu, "A two-step method for missing Spatio-temporal data reconstruction," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 7, p. 187, Jun. 2017.

[32] C. Li, L. J. Chen, X. Chen, M. Zhang, C. P. Pang, and H. Chen, "Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020," *Eurosurveillance*, vol. 25, no. 10, Mar. 2020, Art. no. 2000199.

PEIXIAO WANG is currently pursuing the Ph.D. degree with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research focus on geographic information service and spatio-temporal data mining.

XINYAN ZHU received the Ph.D. degree from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is currently a Professor with Wuhan University. During the past years, he has published over 100 refereed journal articles and conference papers. His research interests cover spatio-temporal ata modeling and analysis, distributed spatial database, holographic position map and its application, indoor location and navigation, social geographic computing, and public health.

WEI GUO received the Ph.D. degree from the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is currently an Associate Professor with Wuhan University. During the past years, he has published over 40 refereed journal articles and conference papers. His research interests cover spatio-temporal ata modeling and analysis, distributed spatial database, and public health.

HUI REN is currently pursuing the M.S. degree with the State Key Laboratory of Information Engineering in Surverying, Mapping and Remote Sensing, Wuhan University. Her research focus on public health and spatiotemporal analysis.

TAO HU received the B.S. and Ph.D. degrees in GIS from Wuhan University. His primary research field is GIScience, with a focus on geospatial data analytics, spatiotemporal analysis with applications to urban crimes, social sensing, and public health. He is currently a Postdoctoral Fellow with the Harvard Center for Geographic since July 2019. He is working on the Spatial Data Lab project. Before that, he was a Postdoctoral Researcher with Kent State University since 2016. He has more than 20 peer-reviewed publications and serves as a peer reviewer for more than ten international journals.

• • •